Faculty Working Papers

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

REVISED ERROR COMPONENTS:
MORE EFFICIENT ESTIMATION WITH COMBINED
CROSS SECTION AND TIME SERIES

Robert W. Resek

#228

1. Introduction

Estimation of combined cross-section time series models is frequently carried out by error components methods as well as by co-variance models. In this paper we analyze the small sample relation between these estimators. In addition we propose a specific new estimator, the revised error components estimator whose small sample efficiency is always greater than either competing estimator.

In this work we rely heavily on previous studies of error component models. Swamy and Arora [10] developed exact finite sample results and an α-class of estimators. As one might expect the revised error component estimator here as well as the r-class which includes it are special cases of their α class. In addition readers will see the close dependence on work of Nerlove [8], Balestra and Nerlove [2], Arora [1], Fuller and Battese [3, 4] and Wallace and Hussain [11].

We proceed by restating briefly known results about error components models. Then we develop in the two component case the r-class of estimators and the specific member which is optimal. Alternative variance estimators are considered and their relative merit determined. Finally, the three component model is developed and the results summarized.

2. Model

Consider a linear econometric model:

(1) $y = \beta_0 i + X\beta + \varepsilon$

i is an (NT x 1) vector of ones. y and ε are (NT x 1). X is (NT x k) and is measured as deviations from column means so that $X'i = 0$. β is (k x 1).

For simplicity there are N elements of the cross section called "states" and T time elements. Each state-time combination has exactly one observation. Within X, y, and ε, the time subscripts vary most rapidly and all time observations within a state are adjacent.

The model (1) will be transformed by premultiplication by appropriate transformation matrices P_1, P_2, etc. where each P has NT columns. These will all be expressed as Kronecker products: e.g.

$$P = A \otimes B$$

where A has N columns and operates on states and B has T columns and operates on time. Now consider the elements of ε. ε_{ij} refers to state i, time j.

$$(2) \quad \varepsilon_{ij} = u_{ij} + v_j + w_i$$

where each u, v, and w is an independent normal random variable and is independent of X. All results are conditional on X in this small sample. The mean of each component is zero and their variances are σ_u^2, σ_v^2, σ_w^2. The normality assumption is necessary with small samples to provide χ^2 and F distributions but we show below the major results do not depend on these distributions. For the analysis we define two new variances:

$$(3) \quad \sigma_1^2 = \sigma_u^2 + T \sigma_v^2$$

$$\sigma_2^2 = \sigma_u^2 + N \sigma_w^2$$

It is clear that $\sigma_1^2 \geq \sigma_u^2$, $\sigma_2^2 > \sigma_u^2$. These inequalities play a very important role below. Employing (2) and (3)

$$E\varepsilon = 0$$

$$(4) \quad V = E\varepsilon\varepsilon' = \sigma_u^2 (I \otimes I) + (\sigma_1^2 - \sigma_u^2)(I \otimes \frac{1}{T} J) + (\sigma_2^2 - \sigma_u^2)(\frac{1}{N} J \otimes I)$$

where $J = ii'$ is an N x N or T x T matrix of ones. The square matrices in (4) are each idempotent.

The BLUE estimator of β is Generalized Least Squares (GLS) but this depends on the unknown matrix V. In turn this matrix depends on three unknown elements σ_u^2, σ_1^2 and σ_2^2. Since V need only be known to a scalar multiple we define:

(5) $\gamma_1 = \sigma_u^2 / \sigma_1^2$

$\gamma_2 = \sigma_u^2 / \sigma_2^2$

Thus we need only know γ_1 and γ_2. During some of the development we will consider the model as stated (the three component model). For simplicity we also consider the two component model in which:

(6) $\sigma_w^2 = 0$, $\sigma_2^2 = \sigma_u^2$, $\gamma_2 = 1$

In this situation there is one unknown value γ_1, which will be called γ.

3. Useful Transformations

The estimation is facilitated by use of various transformations. Consider a T by T Helmert matrix ϕ [9, p. 13]. The first column is ϕ_1 and the remaining (T-1) columns are ϕ_2. Then $\phi_1 = T^{-1/2}i$ and ϕ is orthogonal. Therefore,

$$\phi\phi' = I_T; \quad \phi'\phi = I_T; \quad \phi_1'\phi_1 = 1; \quad \phi_2'\phi_2 = I_{T-1}$$

$$\phi_1'\phi_1 = \frac{1}{T} J; \quad \phi_2'\phi_2 = I - \frac{1}{T} J; \quad \phi_1'\phi_2 = 0$$

Now define four transformations which will be used with the two component model.

(7) $P_1 = I \otimes \phi_1'$ $P_3 = I \otimes \phi_1\phi_1' = I \otimes \frac{1}{T} J = P_1'P_1$

   $P_2 = I \otimes \phi_2'$ $P_4 = I \otimes \phi_2\phi_2' = I \otimes (I - \frac{1}{T} J) = P_2'P_2$

where $P_3 + P_4 = I$; $P_3 P_4 = 0$.

When P_1 and P_2 are vertically augmented they yield the orthogonal matrix $I \otimes \phi$ and so together provide independent data sets which convey all sample information.

   We also define (where ϕ is now N x N).

   $P_5 = \phi_1\phi_1' \otimes I = \frac{1}{N} J \otimes I$

   $P_6 = \phi_2\phi_2' \otimes I = (I - \frac{1}{N} J) \otimes J$

so that

(8) $V = \sigma_u^2 I + (\sigma_1^2 - \sigma_u^2) P_3 + (\sigma_2^2 - \sigma_u^2) P_5$

   We will use this notation to develop initial estimates of β as well as estimates of the unknown variances. We turn exclusively to the two component model until the results in that case are developed.

(9) $V = \sigma_u^2(I - P_3) + \sigma_1^2 P_3 = \sigma_u^2 P_4 + \sigma_1^2 P_3$

(10) $V^{-1} = \frac{1}{\sigma_u^2} P_4 + \frac{1}{\sigma_1^2} P_3$ since $P_3' + P_4 = I$

4. Estimation of σ_u^2 and σ_1^2

   Transform (1) by P_1:

(11) $P_1y = \beta_o P_1 i + P_1 X \beta + P_1 \epsilon$

(12) $V_1 = EP$, $\varepsilon\varepsilon'P_1 = \sigma_1^2 I$

V_1 is N by N and ordinary least squares (OLS) yields BLUE estimates given the transformed data.

Define

(13) $Z_1 = X'P_1'P_1X = X'P_3X$

(14) $b_1 = Z_1^{-1} X'P_1'P_1y$

(15) $b_o = i'P_1y/N = \Sigma y_{ij}/NT = $ Grand Mean

(16)   var $b_1 = \sigma_1^2 Z_1^{-1}$

(17)   var $b_o = \sigma_1^2/NT$

The residual vector e_1 is employed to estimate σ_1^2:

(18) $\hat{\sigma}_1^2 = e_1'e_1/q$

(19) $q = N-1-k$

$\hat{\sigma}_1^2$ follows a Chi Square distribution with q degrees of freedom. Equation (14) implies that we assume Z_1 nonsingular and hence positive definite. The analysis can proceed without this assumption but that generalization always brings unwanted confusion. The reader may develop results without it.

Similarly P_2 may be employed (and we assume Z_2 nonsingular).

(20) $P_2y = P_2X\beta + P_2\varepsilon$

(21) $V_2 = \sigma_u^2 I$ (NT-N) square

$Z_2 = X'P_2'P_2X = X'P_4X$

(22) $b_2 = Z_2^{-1} X' P_2' P_2 y$

(23) $\text{var } b_2 = \sigma_u^2 Z_2^{-1}$

The residuals e_2 are employed to estimate σ_u^2:

(24) $\sigma_u^2 = e_2' e_2 / n$

$n = N(T-1) - k$

As above σ_u^2 follows a chi square distribution.

We note that b_1 and b_2 are independent and include all the sample information. They are independent of $\hat{\sigma}_1^2$ and $\hat{\sigma}_u^2$ which are also pair wise independent. We will rely on an estimate of γ:

(25) $\hat{\gamma} = \hat{\sigma}_u^2 / \hat{\sigma}_1^2$

$\hat{\gamma}$ is clearly distributed F with (n, q) degrees of freedom.

Before proceeding, consider P_3 and P_4.

$P_2' P_2 = P_4' P_4$ so $b_2 = b_4$. Also $e_4 = P_2' e_2$ so that:

$e_4' e_4 = e_2' P_2 P_2' e_2 = e_2' (I \otimes \phi_2' \phi_2^{`}) e_2 = e_2' e_2$

Thus P_2 and P_4 yield identical coefficient vectors and variance estimates. The same can be shown to be true of P_1 and P_3.

P_2 and P_4 are also identical to the direct estimation of (1) by ordinary least squares after (N-1) state dummy variables are added. It is the covariance (CV) estimate and we will refer to b_2 by that name.

For the development we employ only the estimators of the variance but other estimators exist which are in some ways superior and these are

discussed below. We will now develop the Generalized Least Squares and Error Components estimators based on the analysis so far.

5. Generalized Least Squares

The GLS estimator can be written employing this notation.

$$b_g = [X'V^{-1}X]^{-1} [X'V^{-1}y]$$

$$X = P_3 X + P_4 X$$

Employing (10, (13), (22)

$$X'V^{-1}X = \left\{ X'(P_3 + P_4)(\frac{1}{\sigma_u^2}P_4 + \frac{1}{\sigma_1^2}P_3)(P_3 + P_4) X \right\}$$

$$= \left\{ \frac{1}{\sigma_u^2}X'P_4 X + \frac{1}{\sigma_1^2}X'P_3 X \right\}$$

$$= \left\{ \frac{1}{\sigma_u^2}Z_2 + \frac{1}{\sigma_1^2}Z_1 \right\}$$

$$b_g = \left\{ \frac{1}{\sigma_u^2}Z_2 + \frac{1}{\sigma_1^2}Z_1 \right\}^{-1} \left\{ \frac{1}{\sigma_u^2}X'P_4 y + \frac{1}{\sigma_1^2}X'P_3 y \right\}$$

Employing (5), (14), and (23)

(26) $$b_g = [Z_2 + \gamma Z_1]^{-1}[X'P_4 y + \gamma X'P_3 y]$$

(27) $$b_g = W_2^* b_2 + W_1^* b_1$$

where $W_2^* = [Z_2 + \gamma Z_1]^{-1}X'P_4$

$$W_1^* = [Z_2 + \gamma Z_1]^{-1}\gamma X'P_3$$

$$W_2^* + W_1^* = I$$

(28) $\quad \text{var} (b_g) = \sigma_u^2 [Z_2 + \gamma Z_1]^{-1}$

These are the well known results which depend on <u>known</u> values for σ_u^2, and σ_1^2 or γ. (27) shows that b_g is a weighted average of b_2 and b_1.

6. Revised Error Components

We introduce a class of estimators, the r-class, of which one member will be Revised Error Components (REC). For the r class, γ in (26) is replaced by $r\hat{\gamma}$ where $0 \le r$. The r class is a subset of the α class of estimators discussed by Swamy and Arora. However, their purpose was the broad proof of asymptotic distribution, etc. and their multiple parameters, while desirable for that purpose, are not desirable here.

The r class estimator is:

(29) $\quad b_r = [Z_2 + r\hat{\gamma} Z_1]^{-1} [Z_2 b_2 + r\hat{\gamma} Z_1 b_1]$

$$= W_{2r} b_2 + W_{1r} b_1$$

where

$$W_{2r} + W_{1r} = I$$

$$E(b_r | \hat{\gamma}) = W_{2r} \, Eb_2 + W_{1r} \, Eb_1 = (W_{2r} + W_{1r}) \, \beta = \beta$$

Hence b_r is conditionally unbiased for any $\hat{\gamma}$ and r and so also is unconditionally unbiased. The variance of b_r conditional on $\hat{\gamma}$ is

(30) $\quad \text{var}(b_r | \hat{\gamma}) = \sigma_u^2 [Z_2 + r\hat{\gamma} Z_1]^{-1} [Z_2 + (r^2 \hat{\gamma}^2 / \gamma) Z_1] [Z_2 + r\hat{\gamma} Z_1]^{-1}$

Although expressions like (30) are frequently employed to characterize the efficiency of (29), their use is incorrect because $\hat{\gamma}$ is a random variable estimated in the process. The correct matrix is the expectation of (30) taken over $\hat{\gamma}$.

The r class is interesting because it includes other major estimators as special cases. Error components (EC) has $r = 1$. Covariance (CV) has $r = 0$, GLS has $r = 1$ and $\hat{\gamma} = \gamma$. Finally REC which is specified below is a member of this class.

Since every member of the r-class is unbiased we develop REC by minimizing the variances. In order to proceed we need additional assumptions about Z_1 and Z_2. For the development of estimates we employ an overly strict assumption and subsequently relax it. Our final assumptions are more specific than we would desire but we believe that other analysis including Monte Carlo analysis will show the particular estimator to be more efficient than alternatives in any real case. This is because there is substantial breadth to what is employable in our final assumptions.

7. Relation Between Z_1 and Z_2.

For this discussion we consider three different factors entering Z_1 and Z_2. These matrices have different number of underlying observations; those observations have different variation; and the pattern of variation is different.

The first two elements, if considered without the third, lead to our two matrices being proportional to each other. Consider the number of observations: Z_1 derives from P_1 and has N state means, while Z_2 (from P_2) has $N(T-1)$ deviations from those means. Hence Z_2 has $(T-1)$

times the observations of Z_1 and will be (T-1) times as large from this factor above.

Second, we observe that often variation between states is larger (say m times as large) as variation within states, so that each <u>observation</u> making up Z_1 will be m times as large as one in Z_2. Combining these two elements Z_1 is $m/(T-1) = \lambda$ times the size of Z_2.

The third source of variation is the "pattern of variation." Thus two variables in X may be very collinear within states (with large cross product in Z_2), but unrelated between states. One very important reason for employing error components estimation rather than covariance is the full use of this information source. We temporarily restrict ourselves to the first two sources of variation and make this assumption:

<u>A1</u> Z_1 is proportional to Z_2

(31) $Z_1 = \lambda Z_2$

8. Optimization of r and Efficiency Comparison

Rather than looking directly at the variance of b_r, we define the proportion by which each variance matrix element exceeds the corresponding element in the variance of the GLS estimator. Our assumption A1 makes this relative variance (RV) a constant scalar for each matrix. Thus for GLS under A1:

(32) $\text{var}(b_g) = \sigma_u^2 (1 + \lambda\gamma)^{-1} Z_2^{-1}$

For the r class:

(33) $\text{var}(b_r | \hat{\gamma}) = [1 + \text{RV}(b_r | \hat{\gamma})]\, \text{var}(b_g)$

(34) $\quad RV(b_r|\hat{\gamma}) = (1 + \lambda\gamma)(1 + \lambda r^2\hat{\gamma}^2/\gamma)/(1 + \lambda r\hat{\gamma})^2 - 1$

$$= \lambda\gamma(1 - r\hat{\gamma}/\gamma)^2/(1 + \lambda r\hat{\gamma})^2$$

For CV we set $r = 0$ and for EC we set $r = 1$:

(35) $\quad$ CV: $\quad RV(b_2) = \lambda\gamma$

(36) $\quad$ EC: $\quad RV(b_e|\hat{\gamma}) = \lambda\gamma(1 - \hat{\gamma}/\gamma)^2/(1 + \lambda\hat{\gamma}^2)$

Our goal now is to minimize (34) with respect to r. The resulting estimator will have its RV compared with (35) and (36). Clearly we remain interested in the expected value of (34) and (36) over the distribution of $\hat{\gamma}$.

Examine the denominator of (34). Employing the inequality at (3), we have $\sigma_1^2 \geq \sigma_u^2$ or $\gamma \leq 1$. Because the true parameter satisfies this constraint, we will apply it also to the estimate. Thus (25) is replaced by

(37) $\quad \hat{\gamma} \begin{cases} = \hat{\sigma}_u^2/\hat{\sigma}_1^2 & \text{if} \quad \hat{\sigma}_1^2 \geq \hat{\sigma}_u^2 \\ = 1 & \text{if} \quad \hat{\sigma}_1^2 < \hat{\sigma}_u^2 \end{cases}$

This constraint is widely employed in current estimation by Error Components. Given this constraint:

(38) $\quad 1 \leq 1 + \lambda r\hat{\gamma} \leq 1 + \lambda r$

In addition from the discussion of section 7 we know λ to be fairly small in general and below we set $r \leq 1$. Thus the bounds provided by the inequality are relatively narrow. We note however that under relaxed

assumptions below the counterpart of λ may be greater than 1.

Employ (38) with (34):

$$(39) \qquad RV(b_r|\hat{\gamma}) \begin{cases} \leq \lambda\gamma(1 - r\hat{\gamma}/\gamma)^2 \\ \\ \geq \lambda\gamma(1 - r\hat{\gamma}/\gamma)^2/(1 + \lambda r)^2 \end{cases}$$

A similar pair of inequalities holds for EC when we set $r = 1$ in (39).
Rather than minimizing the relative variance over r, we will minimize
each limit as given by (39). Consider now the distribution of $\hat{\gamma}/\gamma$.
Define its mean and variance:

$$(40) \qquad \mu = E(\hat{\gamma}/\gamma)$$

$$\sigma^2 + \mu^2 = E(\hat{\gamma}/\gamma)^2$$

Then take the expectation of (39)

$$(41) \qquad RV(b_r) = E[RV(b_r|\hat{\gamma})] \leq \lambda\gamma[1 - 2r\mu + r^2(\sigma^2 + \mu^2)]$$

$$\geq \lambda\gamma[1 - 2r\mu + r^2(\sigma^2 + \mu^2)]/(1 + \lambda r)^2$$

The optimal r values for the upper and lower limits are:

$$(42) \qquad r_1 = \mu/(\sigma^2 + \mu^2)$$

$$r_2 = (\mu + \lambda)/(\sigma^2 + \mu^2 + \lambda)$$

Since we wish to minimize the largest possible variance, we place greatest
weight on the upper limit value, r_1. Also we note that as λ approaches zero,
r_2 goes to r_1 and that r_1 is free of dependence on λ.

Thus REC, the revised error component estimator is defined as the member of the r class with $r = r_1$. The estimator is called b*. This value in principle can be determined whether or not Al is true and we advocate its use in general. As previously indicated we shall below indicate other cases where it is optimal.

Under assumption Al we find these conclusions concerning EC, CV, and REC:

C1.   REC is more efficient than CV

C2.   REC is known to be more efficient than EC in this sense:

        a.   The upper limit of the inequality (42) for REC is below the upper limit for EC.

        b.   The lower limit of the inequality is below the lower limit for EC as long as $\lambda^2 < \sigma^2 + \mu^2$. Since $\mu \geq 1$ in most cases, this allows a very large λ.

        c.   RV (b*) is strictly less than $RV(b_e)$ by (42) if $(1 + \lambda)^2 < \sigma^2 + \mu^2$. We believe in practice b* will always be superior to b_e.

C3.   EC is more efficient than CV if $\sigma^2 + \mu^2 < 2\mu$. It is less efficient if $\sigma^2 + \mu^2 > 2(\mu + \lambda) + \lambda^2$. Intermediate cases are uncertain.

Following section 10 we provide a specific simple example so one can see the numerical efficiency gain.

9. Distribution of $\hat{\gamma}$.

The preceding results are specific and useful for any distribution of $\hat{\gamma}$. We now turn to the specific distribution that arises from our

estimates (25) and (37). The distribution stated below (25) depends on the original assumption that all errors have normal distribution. In this case $(\hat{\gamma}/\gamma)$ was F with n and q degrees of freedom. For the present we ignore the restriction that $\hat{\gamma} \leq 1$ but return to it later. We recall that n, defined in (24) depends on NT and will in almost every case be fairly large. q on the other hand is N-1-k and can be very small. For example, one could have N = 7, k = 5 so q = 1. Indeed an important reason for later considering other estimates for σ_1^2 is that q here can be zero so the estimate would not exist.

For this F distribution we have (for q > 4)

$\mu = q/(q-2)$

$\sigma^2 + \mu^2 = [q^2(n + 2)]/[(q - 2)(q - 4)n]$

(43) $r_1 = [(q - 4)n]/[q(n + 2)]$

We see that the mean of F becomes infinite when q = 2 and σ^2 does when q = 4.

In (38) we found that the fact that $\hat{\gamma} \leq 1$ was very useful. It plays an extremely important role here, too. Under (37) $\sigma_1^2 > \sigma_u^2$, $0 \leq \hat{\gamma} \leq 1$ so $0 \leq \hat{\gamma}/\gamma \leq 1/\gamma$. Thus the range of $\hat{\gamma}$ is finite and all moments therefore must exist. The distribution is now a "truncated F" $[F_t(n, q, F*)]$ with (n, q) degrees of freedom and an upper limit F*. By this we mean the corresponding F for values $0 \leq F_t < F*$. There is a "spike" of F* such that $P(F_t = F*) = P(F \geq F*)$. In our situation $F* = 1/\gamma$. Since γ is an unknown parameter we cannot know F* or the resulting mean or variance of F_t but can examine them conditional on various F*.

The moments of F_t (n, q, F*) were found by numerical integration and appear in Table 1. $n \to \infty$ was employed since in general use, n will be large. The truncation reduces μ and $(\sigma^2 + \mu^2)$ and also increases r. It is clear that a prior point estimate or distribution of γ and hence F* is required to select the most efficient estimator and the reader may consider the optimal use of such prior information. We propose that a broad rule of thumb be employed and will show where it is better than alternatives. The rule is

(44) $r = \quad [(q + 4)n]/[(q + 11)(n + 2)] \quad$ if $q \leq 15$

$\quad\quad\quad\quad [(q - 4)n]/[q(n + 2)] \quad\quad\quad$ if $q > 15$

This rule was devised as a simple way to approximate Table 1 when F* = 3 and to include the effect of n. It may be shown that it leads to a smaller relative variance than EC for the values in the Table as long as either $F* \geq 5$ or $q \geq 4$. REC is now defined as the member of the r-class with r defined as in (44). Two conclusions can be derived.

C4.   REC is superior to either EC or CV if $F* \leq 5$ or $q \geq 4$. If $F* > 5$ and $q \leq 3$ then the r-class should be employed with appropriate (small) value of r.

C5.   EC is superior to CV if: F* = 2 for any q; F* = 3 and $q \geq 4$; F* = 5, and $q \geq 7$; $F* > 5$ and $q \geq 8$. This is found by direct application of C3 to Table 1. Clearly the small sample properties are very dependent on the true unknown γ. Monte Carlo studies should reach various conclusions according to the variances selected in advance for analysis.

It is important to note that in Table 1, the biggest effect on r is created by F* and not the degrees of freedom. This implies intuitively

Table 1

Truncated F Distribution
Mean, Variance, Optimal r
Numerator Degrees of Freedom → ∞

Denominator Degrees of Freedom, q		Truncation = F*				
		2.	3.	5.	10.	∞
1.	μ	1.44	1.92	2.69	4.14	*
	$\sigma^2 + \mu^2$	2.54	4.90	11.01	32.35	*
	r	.57	.39	.24	.13	0
2.	μ	1.35	1.68	2.13	2.78	*
	$\sigma^2 + \mu^2$	2.23	3.87	7.41	16.82	*
	r	.60	.43	.29	.16	0
3.	μ	1.30	1.55	1.83	2.16	3.00
	$\sigma^2 + \mu^2$	2.05	3.29	5.52	10.10	*
	r	.63	.47	.33	.21	0
4.	μ	1.26	1.46	1.65	1.82	2.00
	$\sigma^2 + \mu^2$	1.93	2.89	4.36	6.72	*
	r	.65	.51	.38	.27	0
5.	μ	1.24	1.39	1.52	1.62	1.67
	$\sigma^2 + \mu^2$	1.84	2.60	3.59	4.87	8.33
	r	.67	.53	.42	.33	.20
6.	μ	1.22	1.34	1.43	1.49	1.50
	$\sigma^2 + \mu^2$	1.77	2.38	3.05	3.78	4.50
	r	.69	.56	.47	.39	.33
7.	μ	1.20	1.30	1.37	1.40	1.40
	$\sigma^2 + \mu^2$	1.71	2.20	2.67	3.10	3.27
	r	.70	.59	.51	.45	.42
8.	μ	1.19	1.27	1.32	1.33	1.33
	$\sigma^2 + \mu^2$	1.65	2.06	2.39	2.66	2.67
	r	.72	.62	.55	.50	.50

$$r = \mu/(\sigma^2 + \mu^2)$$

* = infinite

that the results will not materially be altered if the underlying distribution is different from F as long as the truncation is present.

We have at this point developed the REC estimator and have shown it to be broadly superior to EC and to CV. It remains to expand the analysis in three ways. These are first to relax the initial assumption A1; second to consider other estimators of σ_1^2 and σ_u^2; and third to present the estimators for a three component model.

10. Alternative Assumptions about Z.

The analysis above depends on A1 and we now discuss a variety of means by which comparable results may be achieved. While the family ultimately outlined here is not all inclusive, it contains a wide family of different kinds of situations. Initially we will state the major alternative assumption and then, since its meaning may not be instantly clear we will consider specific situations under which it will arise.

For the matrices Z_1 or Z_2 we may write eigen vectors and eigen values in a matrix equation; e.g.

$$Z_2 A_2 = A_2 \Lambda_2$$

where A_2 is an orthogonal matrix of eigen vectors and Λ_2 a diagonal matrix of eigen values. A_2 may not be unique because its columns may be permuted along with elements of Λ_2 and if there are multiple eigen values, there are multiple solutions. We assume the following:

A2   For the matrices Z_1 and Z_2 there exist matrices of eigen vectors A_1 and A_2 such that $A_1 = A_2 = A$. It is permissible to perform a single variance transformation on both Z matrices before finding eigen vectors.

One way of achieving this is assumption A1 above. Another direct way is the following:

__A3__ Z_1 and Z_2 are diagonal matrices. This would arise when all the Xs are uncorrelated both within states and between states. Incidentally, this clearly includes the case where there is a single regressor.

Other possibilities are best approached by first performing a variance transformation on Z_1 and Z_2 as mentioned in the assumption. Form the diagonal matrix Q such that each diagonal element is the square root of the reciprocal of the diagonal of Z_2, i.e.

(45) $$Q_{ii} = Z_{2ii}^{-1/2}$$

$$Q_{ij} = 0 \qquad\qquad i \neq j$$

Define

(46) $$Z_2^* = QZ_2Q$$

$$Z_1^* = QZ_1Q$$

It is clear that by construction, the diagonal of Z_2^* consists of ones. We assume:

__A4__   The diagonal elements of Z_1^* are equal to each other. The implication of this is that variances in Z_1 are proportional to those in Z_2. This transformation will be made before A2 is applied. Its use, as seen below, is to transform both X and β without change of the content of the analysis. Along with A4, anyone of the following is sufficient for A2.

<u>A5</u>   There are two explanatory variables in the model (k = 2).

<u>A6</u> Z_1^* and Z_2^* have a single constant every where off the main diagonal (but these two may be different).

<u>A7</u> Z_1^* and Z_2^* are each circular symmetric matrices.

<u>A8</u> Z_1^* and Z_2^* are each tridiagonal matrices.

See Press [9] for specific discussion of the latter two assumptions. Another possibility is that Z_1 and Z_2 are each conformably block diagonal with separate blocks satisfying different assumptions above. In general terms these assumptions say that one makes a single variance transformation to Z_1 and Z_2 and the results are pattern matrices of a single type but different coefficients.

Now we proceed to assume A4 and A2 are consecutively employed.

(47) $$Z_2^* A = (QZ_2Q) A = A\Lambda_2$$
$$Z_1^* A = (QZ_1Q) A = A\Lambda_1$$

Define M as the reciprocal square root of Λ_2 so that:

(48) $$M^2 \Lambda_2 = I$$

Reconsider the original model (1).

(49) $$Y = \beta_0 i + X\beta + \epsilon$$
$$= \beta_0 i + XQAM(QAM)^{-1} \beta + \epsilon$$

Define

(50) $$\delta = (QAM)^{-1}\beta \qquad \text{so that}$$

(51) $$Y = \beta_0 i + (XQAM)\delta + \epsilon$$

We will now examine the variance matrix of the transformed coefficient vector δ. This will be minimized in the sense that alternative estimators will have variance matrices which differ by a positive semi-definite matrix indicating the results will also hold for estimates of β.

The transformation QAM alters the matrices Z_1 and Z_2. These become:

(52) $Z_2^{**} = M'A'Q'Z_2QAM = I$

$Z_1^{**} = M'A'Q'Z_1QAM = M\Lambda_1 M = \Lambda_1 \Lambda_2^{-1} = \Lambda$

Λ is a diagonal matrix where the i th diagonal element is the ratio of the eigen value of Z_1 to the corresponding eigen value of Z_2.

(53) $\lambda_i = \lambda_{1i}/\lambda_{2i}$

We proceed directly to the variance of b_r (30) where b_r is now an estimate of δ.

(54) $\text{var}(b_r|\hat{\gamma}) = \sigma_u^2 [I + r\hat{\gamma}\Lambda]^{-1}[I + (r^2\hat{\gamma}^2/\gamma)\Lambda][I + r\hat{\gamma}\Lambda]^{-1}$

Clearly all off diagonal elements (covariances) are zero and diagonal elements may be written:

(55) $\text{var}(b_{ri}|\hat{\gamma}) = \sigma_u^2[1 + \lambda_i r\hat{\gamma}]^{-2}[1 + (r^2\hat{\gamma}^2/\gamma)\lambda_i]$

Previously this depended on a scalar λ for the entire matrix but now it depends on λ_i. Relative variance continues to be as in (33) but is unique for each parameter in δ.

(56) $\text{RV}(b_{ri}|\hat{\gamma}) = \lambda_i\gamma (1-r\hat{\gamma}/\gamma)^2/(1 + \lambda_i r\hat{\gamma})^2$

This result is identical to (34) except that the subscript is added to λ. Our justification for choosing the upper limit in (42) to determine r is strengthened because we are free of dependence on λ. All subsequent results hold in their entirety but now apply to each parameter being estimated. Cl can be restated as follows:

The variance of the CV estimator of δ exceeds that of the REC estimator by a positive semi-definite matrix. Since this difference is true for estimator of δ, it is also true of the comparable estimators of β. Each other conclusion has a comparable interpretation showing in particular REC to be better than the alternatives. We should note that in some cases λ_1 may be fairly large leading to wide inequalities in (42) but also leading larger relative efficiency differences.

An example is presented to make the potential efficiency gain more real. This example falls under A4 and A5.

$$X'P_2X = Z_2 = \begin{bmatrix} 16 & -36 \\ -36 & 100 \end{bmatrix} \qquad X'P_3X = Z_1 = \begin{bmatrix} 4 & 8 \\ 8 & 25 \end{bmatrix}$$

The only restriction is that variances are proportional: i.e. $16/100 = 4/25$. The example was selected so the pattern of variation differs substantially, that is, the cross products are large negative and positive numbers (-36 and $+8$). The variance transformation is performed and then eigen vectors found:

$$Q = \begin{bmatrix} .25 & 0 \\ 0 & .10 \end{bmatrix} \qquad A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} .$$

The eigen values are:

$$\lambda_{21} = [\ .1 \quad 1.9] \qquad \lambda_{11} = [.45 \quad .05]$$

$$\lambda_1 = 4.5 \quad \lambda_2 = .02631$$

Assume also $\gamma = 1/3 \quad q = 5 \quad n = 98$

Then by (44) $r = .551$

For δ_1 $RV(CV) = 1.5$

           $RV(EC) \leq 1.5 \ (.82) = 1.23$

           $RV(REC) \leq 1.5 \ (.26) = .39$

For δ_2 $RV(CV) = .0088$

           $RV(EC) \leq .0072$

           $RV(REC) \leq .0022$

      The figures indicate that there is little difference in the three estimators relative to δ_2 but substantial difference in δ_1. For δ_1, CV has variance that is 150% larger than GLS with known variance matrix while EC has variance up to 123% larger and REC up to 39% larger. Thus the variance of CV is 12% greater than that of EC and 80% greater than REC. For δ_2, the ordering is the same but REC is only known to be 0.66% superior to CV.

      This concludes the discussion of alternative assumptions about Z. We turn in the next section to alternative variance estimates.

11. Alternative Variance Estimates.

The estimates above of σ_1^2 and σ_u^2 (18) and (23) are not the only such estimates. Arora [1] employs those given here. Fuller and Battese [3, 4] use $\hat{\sigma}_u^2$ but for σ_1^2 employs the estimator below, while Wallace and Hussain [11] use estimates which are each different. We will consider the merits of the major alternative forms.

Recall that $\hat{\sigma}_1^2$ and $\hat{\sigma}_u^2$ are advantageous because they are each chi square with q and n degrees of freedom respectively. Their disadvantage which we will show, is that they "lose " more degrees of freedom than the model requires and hence are, in certain cases, inefficient. This is of importance particularly when there are few (or zero) degrees of freedom. In the following we need some known results concerning quadratic forms.

1. Assume X is $N(0,\sigma^2 I)$ and A is idempotent of rank r.

Then

$$Q = X'AX \quad \sim \text{ Chi square } (r)$$

Its cumulants are [6, p 168]

(57) $\kappa_s(Q) = C_s \cdot r$

where $C_s = 2^{s-1} (s-1)!$

The cumulants convey the same information as moments [5, p.20] The first four are:

$$\kappa_1 = \mu_1'; \; \kappa_2 = \sigma^2; \; \kappa_3 = \mu_3; \; \kappa_4 = \mu_4 - \sigma^4$$

Where μ_3 and μ_4 are the third and fourth moments about the mean. All cumulants for $s \geq 3$ in a normal distribution are zero.

2. Assume $X \sim N(0,V)$ and A is a general symmetric matrix. Then

$Q = x'Ax$ has the following cumulants [7, p. 153].

(58) $\kappa_s(Q) = C_s \Sigma \lambda_1^s = C_s \text{ tr } (AV)^s$

where the λ_1 are the eigen values of AV and tr (AV) is its trace.

3. Assume $X \sim N(0,V)$ and A and B are general quadratic forms. Then

$$\text{covar}(X'AX, X'BX) = 2 \text{ tr } (AVBV).$$

This is obvious from the variance of $X'(A + B)X$.

Now we will examine variance estimates of the type employed by Wallace and Hussain [11]. First employ ordinary least squares to estimate (1) and find the residual vector e.

$$e = M\varepsilon = [I - (\frac{1}{n}J \otimes \frac{1}{t}J) - X(X'X)^{-1}X]\varepsilon$$

Define

$$Q_3^o = e'P_3 e = \varepsilon'MP_3 M\varepsilon$$

$$Q_4^o = e'P_4 e = \varepsilon'MP_4 M\varepsilon$$

The distribution of Q_3^o and Q_4^o depend on

$$Q_3 = P_3 MVMP_3$$

$$Q_4 = P_4 MVMP_4$$

Where these are symmetric matrices which serve the role of AV in (58). First examine Q_4:

$$Q_4 = \sigma_u^2[P_4 - K_2^o] + (\sigma_1^2 - \sigma_u^2)K_{34}^o$$

where
$$K_i^o = P_i X(X'X)^{-1} X' P_i$$

$$K_i = (X'X)^{-1} X' P_i X$$

$$k_i = \text{tr } K_i = \text{tr } K_i^o$$

$$K_{ij} = K_i K_j \qquad \text{etc. for any sequence of subscripts}$$

$$k_{ij} = \text{tr } K_{ij}$$

$k_{i \ j}^{\ r \ s}$ is k with r subscripts each equal to i followed by s each equal to j.

$$K_i = (Z_1 + Z_2)^{-1} Z_i \qquad \text{and is non negative definite. Hence}$$

$$k_i \geq 0$$

$$k_3 + k_4 = k$$

$$k_{33} + k_{34} = k_3 \qquad \text{etc. for any complementary pair.}$$

Employing eigen values one may show

$$k_{33} \geq k_3^{\ 2}/k$$

$$k_{34} \leq k_3 k_4 / k$$

We may proceed further under the assumptions employed above.

Assume <u>A1</u>:

$$k_3^{\ r} {}_4^{\ p} = \mu^r (1 - \mu)^p k \qquad \mu = \lambda/(1 + \lambda)$$

Assume <u>A2</u>:

$$k_3^{\ r} {}_4^{\ p} = \sum_i \mu_i^{\ r} (1 - \mu_i)^p \qquad \mu_i = \lambda_i/(1 + \lambda_i)$$

12. Estimate of σ_u^2.

Define

(59) $\qquad \sigma*_u^2 = Q_4^{\;o}/m_4$

(60) $\qquad m_4 = N(T - 1) - k_4$

(61) $\qquad E\sigma*_u^2 = \sigma_u^2(1 + yk_{34}/m_4)$

(62) $\qquad y = (\sigma_1^2/\sigma_u^2) - 1$

(63) $\qquad \text{var } \sigma*_u^2 = 2\sigma_u^4 \frac{1}{m_4} [1 - \frac{1}{m_4} \{k_{34} - 2yk_{343} - y^2 k_{3433}\}]$

(61) and (63) show the dependence of this estimators on y and therefore on σ_1^2/σ_u^2. Hence we must assume y is small. Given this and the small size of k_{34} we note that $\sigma*_u^2$ is approximately unbiased and the last term in the variance equation is 1. If y is small the cumulants are:

(64) $\qquad \kappa_s(\sigma*_u^2) = C_s \sigma_u^{2s} m_4^{1-s} [1 - \frac{1}{m_4}(k - k_3^{\;s}) + O(k_{34}/m_4)]$

The last term signifies terms of order at most k_{34}/m_4. Deleting that term

(65) $\qquad \kappa_s(\sigma*_u^2) \leq C_s \sigma_u^{2s} m_4^{1-s}$

which are the cumulants of chi square with m_4 degrees of freedom. Hence $\sigma*_u^2$ is superior, in the sense of cumulants, to such a distribution and hence also to $\hat{\sigma}_u^2$.

The smallness of y can be made more specific by employing A2. In that case

(66) $\qquad \kappa_s(\sigma*_u^2) = C_s \sigma_u^{2s} m_4^{1-s} [1 + \frac{1}{m_4}\{-k + \sum_i^k \mu_i^{\;s}(1 + y - y\mu_i)^s\}]$

This is superior to $\chi^2(m_4)$ if the term in { } is negative for <u>all</u> s. In turn this is true if each element in the sum is less than or equal to 1 or:

$$1 > \mu_i^s(1 + y - y\mu_i)^s$$

$$y \leq 1/\mu_i$$

This is also a necessary condition. In our example above

$$\lambda_i = (4.5, .026) \text{ so } \mu_i = (.82, .075)$$

thus

$$y \leq 1.22 \quad \text{or} \quad \sigma_1^2 \leq 2.22 \, \sigma_u^2 \quad \text{or} \quad \gamma \geq .45.$$

In every case under <u>A2</u>, $\sigma_1^2 < 2\sigma_u^2$ will suffice so that all cumulants of $\sigma*_u^2$ are less than those of $\hat{\sigma}_u^2$.

In many cases under <u>A1</u> or <u>A2</u>, the largest λ will be, say 1/10. In this case $\mu = 1/11$ and $y \leq 11$ is sufficient for $\sigma*_u^2$ to be superior to $\hat{\sigma}_u^2$. Therefore, we conclude that $\sigma*_u^2$ is the superior estimator in cumulant if either σ_1^2/σ_u^2 is small or the largest root λ is small.

Despite the advantages of $\sigma*_u^2$, the gain will in almost every case be small because there are many degrees of freedom. Hence it seems wise to generally employ $\hat{\sigma}_u^2$.

13. Estimates of σ_1^2.

$$Q_3 = \sigma_1^2 \left[(I - \tfrac{1}{N}J) \otimes \tfrac{1}{T} J - 2K_3^{\,o} + K_{33}^{\,o} \right] + \sigma_u^2 \left[K_{33}^{\,o} - K_3^{\,o} \right]$$

Define

(67) $\sigma_1^{*2} = [Q_3 - \hat{\sigma}_u^2 k_{34}]/m_3$

(68) $m_3 = N - 1 - k + k_{44}$

This is the estimator of Fuller and Battese. For simplicity one could employ $Q_s/(N-1)$ or Q_s/m_3 but these require assumptions that $\hat{\sigma}_u^2$ are small, etc. Cumulants can be found generally:

(69) $\kappa_s(\sigma_1^{*2}) = C_s \sigma_1^{2s} m_3^{1-s}[1 - \frac{1}{m_3}(k_{44} - k_4^{2s}) + O(\gamma k_{34}/m_3)]$.

The last term depends on m_3 which may not be very small but also has $\gamma = \sigma_u^2/\sigma_1^2$ which is always less than one. (69) includes the effect of $\hat{\sigma}_u^2$. To make (69) clearer we now employ A2. Under it:

(70) $\kappa_s(\sigma_1^{*2})$

$= C_s \sigma_1^{2s} m_3^{1-s}[1 - \frac{1}{m_3} \sum_i (1 - \mu_i)^2 \{1 - (1 - \mu_i)^{s-2}(1 - \mu_i + \gamma\mu_i)^s\}$

$+ \frac{1}{m_3} (N(T - 1) - k)^{-s}(-\gamma k_{34})^s]$

(71) $\kappa_s(\sigma_1^{*2}) \leq C_s \sigma_1^{2s} m_3^{1-s}$.

The inequality in (71) is true for $s \geq 3$ without any additional assumptions. When $s = 2$, the requirement is that γ be a small amount less than one. One form of requirement for (71) to hold is:

$2[N(T - 1) - k] \geq k/(1 - \gamma)^{1/2}$

For example if $\gamma = .99$ and $k = 4$, then (71) is true if $N(T -1) \geq 24$.

This depends on the fact that at least one $\mu_1 \leq 1/2$. Our real interest is var $(\hat{\sigma}_1^2)$ and one can show

$$\text{var } \sigma_1^{*2} \leq \text{var } (\hat{\sigma}_1^2)$$

if $N(T - 1) - k > 1/4 \, k$

which will of course always be true.

We have shown that σ_1^{*2} is cumulant superior to chi square m_3 as well as to $\hat{\sigma}_1^2$ and hence should always be employed. Clearly there will be some cases with small N where σ_1^{*2} exists and the alternative does not.

We will employ the chi square distribution as an approximation as we return to REC but we note that the truncation of the F distribution means the exact distribution is not required in any event. Thus we propose that REC be employed as above but with $\gamma^* = \hat{\sigma}_u / \sigma_1^{*2}$ replacing $\hat{\gamma}$. The new degrees of freedom are used in defining r. (m_3 replaced q.)

In this section we have shown that the estimates of $\hat{\sigma}_1^2$ of Fuller and Battese are superior to those of Arora but that both serve well in the REC estimator.

14. Three Component REC Model

Now consider the original model with variance matrix given by (4). The analysis was developed with $\sigma_2^2 = \sigma_u^2$ but that assumption is now dropped and the third component restored. We employ four transformations defined as in section 3. These are

(72) $P_{35} = P_3 \cdot P_5$ $P_{36} = P_3 \cdot P_6$

     $P_{45} = P_4 \cdot P_5$ $P_{46} = P_4 \cdot P_6$

One may show

$$P_{34} + P_{45} = P_5; \quad P_{36} + P_{46} = _6; \quad P_{35} + P_{36} = P_3; \text{ etc.}$$

When the model is transformed by P_{35} we have the grand mean for all observations - which is zero within X. P_{45} yields mean values for each time period; P_{36} yields mean values for each state and P_{46} yields deviations from both the state mean and time mean.

When regressions are performed on each transformed model in turn:

P_{35} yields only the estimate of β_o

P_{46} yields b_4 and a new $\hat{\sigma}_u^2$ which has $[(N - 1)(T - 1) - k]$ degrees of freedom.

P_{36} yields b_1 and $\hat{\sigma}_1^2$ which are identical to b_1 and $\hat{\sigma}_1^2$ above.

P_{45} yields b_5 and $\hat{\sigma}_2^2$ where the latter is Chi Square with $T - 1 - k$ degrees of freedom.

We also could define σ_{*1}^2 and σ_{*2}^2 as was done in the preceeding section and which would have the same implications. All of the estimators under consideration are now weighted averages of b_1, b_4, and b_5:

Define the following.

$$\hat{\gamma}_1 = \hat{\sigma}_u^2/\hat{\sigma}_1^2 \qquad\qquad \hat{\gamma}_2 = \hat{\sigma}_u^2/\hat{\sigma}_2^2$$

$$Z_4 = X'P_4P_5X \qquad\qquad Z_5 = X'P_4P_5X$$

$$Z_1 = X'P_3P_6X = X'P_3X$$

$$Z_4 + Z_5 = Z_2 \,\cdot$$

$$W_o = Z_4 + r_a\hat{\gamma}_1Z_1 + r_b\hat{\gamma}_2Z_5$$

$$W_4 = W_o^{-1}Z_4$$

$$W_1 = r_a\hat{\gamma}_1W_o^{-1}Z_1$$

$$W_5 = r_b\hat{\gamma}_2W_o^{-1}Z_5$$

Then the r class estimator is defined as:

(73) $\qquad b_r = W_4b_4 + W_1b_1 + W_5b_5$

This estimator is again identical to the α class of Swamy and Arora. Indeed the entire 7 parameter α class can be expressed as part of the two parameter r class if we set:

(74) $\qquad r_a = \alpha_o/[\alpha_6(\alpha_1 + \alpha_2\hat{\gamma}_1)]$

$\qquad\qquad r_b = \alpha_3/[\alpha_6(\alpha_4 + \alpha_5\hat{\gamma}_2)]$

Although the α notation was desirable for the purposes of that article, the r notation is simpler and more sensible for general estimation purposes.

There are many specific estimators available for a 3 component model such as this one. For each of the two ratios γ_1 and γ_2 we may employ EC, CV, REC, or omit the component and return to a 2 component model. These imply that $r_a = 1$, $r_a = 0$, $r_a = r*$, or $r_a \hat{\gamma}_1 = 1$. Since the same four options are available for the two components there are $4 \times 4 = 16$ different specifications clearly available.

The variance of b_r is:

$$(75) \qquad \text{var}(b_r | \hat{\gamma}_1, \hat{\gamma}_2) = \sigma_u^2 W_o^{-1} [Z_4 + (r_a^2 \hat{\gamma}_1^2 / \gamma_1) Z_1 + (r_b^2 \hat{\gamma}_2) Z_5] W_o^{-1}$$

We now proceed directly to an assumption about Z similar to that made above:

<u>A9</u>   After a single variance transformation Q is applied to Z_5, Z_1 and Z_4 such that $Z_5^* = Q Z_5 Q$ etc. there exist eigen value matrices for the transformed Z that are identical.

This is applied as above to transform the parameter vector to a new vector δ. The eigen values are contained in Λ_4, Λ_1 and Λ_5.

Define

$$\lambda_{ai} = \lambda_{1i}/\lambda_{4ij} \qquad \lambda_{bi} = \lambda_{5i}/\lambda_{4i}$$

Then the relative variance of b_{ri} may be shown to be:

$$(76) \qquad RV(b_{ri} | \hat{\gamma}_1, \hat{\gamma}_2) = Q^*/(1 + \lambda_{ai} r_a \hat{\gamma}_1 + \lambda_{bi} r_a \hat{\gamma}_2)^2$$

$$\leq Q^*$$

$$Q^* = \lambda_{ai} \gamma_1 (1 - r_a \hat{\gamma}_1/\hat{\gamma}_1)^2 + \lambda_{bi} \gamma_2 (1 - r_a \hat{\gamma}_2/\gamma_2)^2$$

$$+ \lambda_{ai} \lambda_{bi} \gamma_1 \gamma_2 (r_a \hat{\gamma}_1/\gamma_1 - r_b \hat{\gamma}_2/\gamma_2)^2$$

μ_1 and σ_1^2 are the mean and variance of $(\hat{\gamma}_1/\gamma_1)$; μ_2 and σ_2^2 are the mean and variance of $(\hat{\gamma}_2/\gamma_2)$; and σ_{12} is the covariance. If $\hat{\sigma}_u^2$, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are respectively chi square with n^o, q_1, and q_2 degrees of freedom then $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are each F with common numerator but independent denominator. If these distributions are not truncated then:

$$(77) \qquad \sigma_{12} = (2/n^o) \; \mu_1 \mu_2$$

This value will be small when n^o is large.

We minimize Q^* over r_a and r_b. The optimum value for r_a^* after some computation is found to be:

$$(78) \qquad r_a^* = r_1 \left\{ 1 - \frac{\lambda_{bi} \, \gamma_2}{1 + \lambda_{bi}\gamma_2} \left[1 - r_b^* \, (\mu_2 + \sigma_{12}/\mu_1) \right] \right\}$$

r_1 is the optimal r value for the two component model defined in (42). A similar equation holds for r_b^*. It is seen from this that r_a^* is approximated by r_1 if $(\lambda_{bi} \, \gamma_2)$ is small or if q_2 is not small (so that r_b^* is approximately one. Because this approximation exists we _define_ the three component REC estimator as the r class with $r_a = r_1$ and $r_b = r_2$.

It remains for us to contrast the efficiency of the alternative estimators with the three component model. We indicated below (74) the variety of estimation methods available. The selection of estimation method depends on the true value of γ_1 or γ_2. Hence the most efficient method is 2 components if γ is close to one. CV is adviseable if q is very small or λ large and EC is good if q is large and γ not close to one.

The most efficient estimator in every case is REC with appropriately chosen r. In general use one may choose r by the rules given in (42) but

shift to another estimator is special cases indicated by the previous

paragraph. The general use of any other estimator is clearly inefficient.

We have developed the r class of estimators and have shown that a

specific member, REC, should be used in essentially all error component

problems. In the process we have developed substantial information about

small sample properties of error components, covariance, and REC estimators.

Further theoretical and Monte Carlo analysis should broaden the specific

assumptions and results.

References

[1] Arora, S. S., "Error Components Regression Models and Their Applications", Annals of Economic and Social Measurement, 2, October, 1973, 451-61.

[2] Balestra, P. and N. Nerlove, "Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas". Econometrics, 34, July, 1966, 585-612.

[3] Fuller, W. A. and G. E. Battese, "Transformations for Estimation of Linear Models and Nested-Error Structure", Journal of the American Statistical Association, 68, September, 1973, 626-32.

[4] Fuller, W. A. and G. E. Battese, "Estimation of Linear Models with Crossed-Error Structure". Journal of Econometrics, 2, March, 1974.

[5] Johnson, Norman L. and Samuel Kotz, Discrete Distributions, Boston, 1969.

[6] Johnson, Norman L. and Samuel Katz, Continuous Univariate Distributions - 1, Boston, 1970.

[7] Johnson, Norman L. and Samuel Katz, Continuous Univariate Distributions - 2, Boston, 1970.

[8] Nerlove, M., "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections", Econometrica, 39, March, 1971, 359-82.

[9] Press, S. James, Applied Multivariate Analysis, New York, 1972.

[10] Swamy, P.A.V.B., and S. S. Arora, "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models", Econometrica, 40, March, 1972, 253-60.

[11] Wallace, T. D. and A. Hussain, "The Use of Error Components Models in Combining Cross Section with Time Series Data", Econometrica, 37, January, 1969.